MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 3, 2015/2016

DPS5018 – INTRODUCTION TO PROBABILITY AND STATISTICS

(A71 – A78, B176 – B195)

31 MAY 2016
2.30 p.m. – 4.30 p.m.
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **5** pages with **5** questions. Key formulae are given in the Appendix.

2. Answer **ALL** questions.

3. Write your answers in the answer booklet provided.

4. All necessary working steps must be shown.

Question 1

a. Statistics is a field of study that consists of four processes in order to make decision. List out the **FOUR** processes. (4 marks)

b. Identify the population and sample in each statement below: (6 marks)

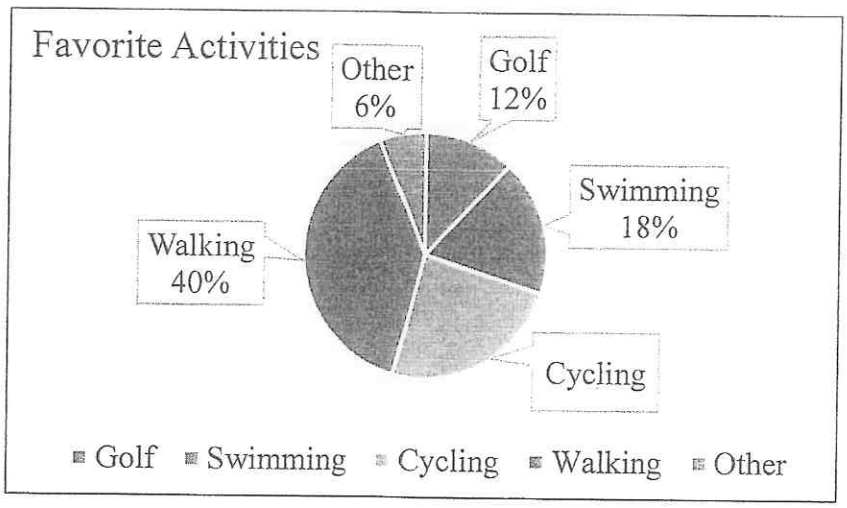
No	Statement	Population	Sample
Eg:	**A list of all asthmatic school children treated in Bukit Beruang clinics in Melaka.**	**All asthmatic school children in Melaka**	**Asthmatic school children treated in Bukit Beruang clinics**
i.	Mr Said conducts a statistical analysis on students' final examination scores in Business Mathematic subject for the past year which is taken by Accounting, Business and IT students. Then, he collects the data only from Accounting students.		
ii.	A group of students from Art Class conduct a survey among their classmates to determine the number of pets owned by the students. The group plan to generalize their result to all students in their school.		
iii.	Mr Chan wants to find the variance of the heights of male students in Algebra class.		

[TOTAL 10 MARKS]

Continued.....

Question 2

a. Pie chart below shows the favorite activities during weekend of 200 students in Class Alpha.



i. Construct a frequency and relative frequency distribution to represent the above data. (3 marks)

ii. If the student ratio in Class Alpha given such male:female are 1:4, construct a table to show the number of students for each gender in each activity. (Round the value to the nearest whole number.) (7 marks)

b. A survey has been done on 40 visitors who registered with Dynamica Gymnasium. The following stem and leaf diagram recorded the number of times that they visited the gymnasium.

Stems	Leaves
3	0 2 3
4	1 3 4 5 6
5	2 4 5 6 7 8 8 8 9 9 9
6	1 3 4 5 6 6 7 8
7	0 2 2 6 7 9
8	3 4 4 5
9	0 1 2

Key: 3 | 0 = 30

i. Based on the stem and leaf diagram above, construct a grouped–data table where the first class starts at 30⩽40. (4 marks)

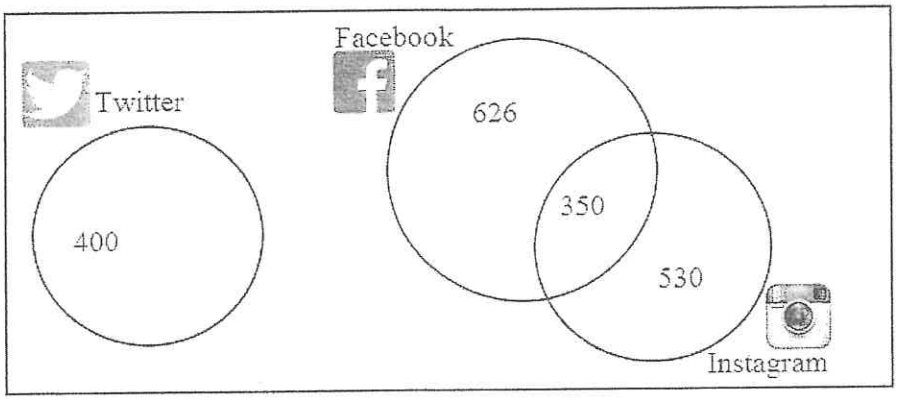
ii. From table in part (i), find the **mean** and **standard deviation.** Round your final answer in three decimal places. (6 marks)

[TOTAL 20 MARKS]

Continued.....

Question 3

The Venn diagram below represents a survey conducted over 2000 students on their preference of social media network.



a. Find the number of students who does not have any of the social media account. (1 mark)

b. Determine two events that are mutually exclusive. (2 marks)

c. If a student is selected at random, find the probability that a student has
   i. Facebook account. (2 marks)
   ii. Instagram account. (2 marks)
   iii. Instagram given that the student has Facebook account. (5 marks)
   iv. either Twitter or Facebook account. (3 marks)

[TOTAL 15 MARKS]

Question 4

a. Let X represent the number of rainy days in Melaka.

X	2	3	4	5	6
$P(X=x)$	$\frac{1}{3}$	Q	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{8}$

   i. Find the value of Q. (3 marks)

   ii. If the average is $\frac{29}{8}$, calculate the variance. (3 marks)

   iii. Compute $P(3 \le X \le 6)$. (3 marks)

b. According to a researcher, there is about 51% chance that a student will buy a textbook. Suppose that five students are selected at random,
   i. find the probability that at most one student will buy a textbook. (3 marks)
   ii. how many of them could we expect would buy the textbook? (2 marks)

Continued.....

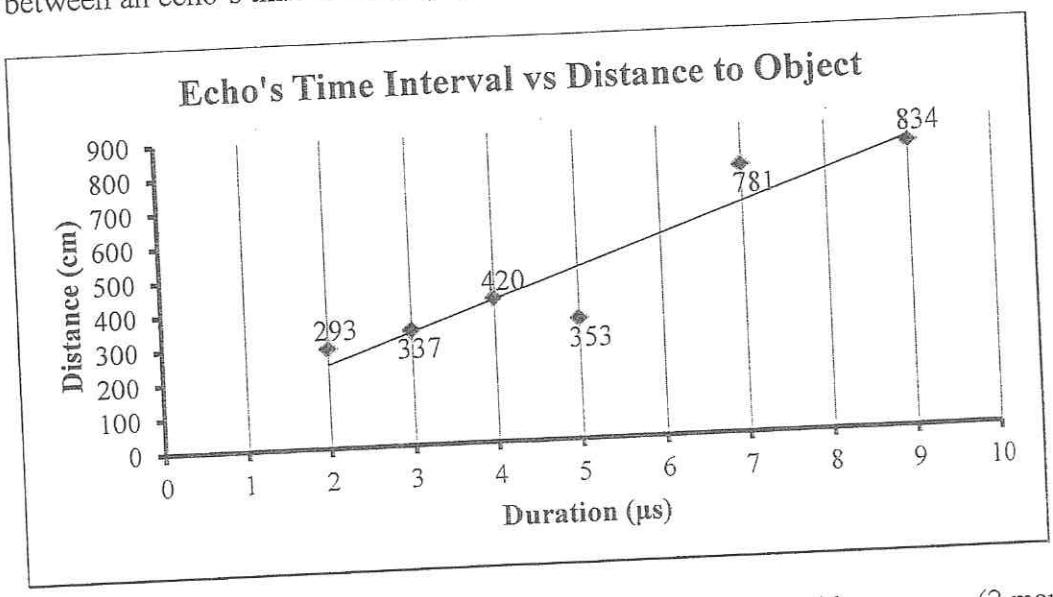
c. Assume that the average number of summons issued by MMU Security Department per week is four which follows a Poisson distribution,
   i. find the probability that no summon is issued in a week. (3 marks)
   ii. what is the percentage of less than five summons were being issued in a week? (3 marks)

d. A group of new students' CGPA is normally distributed with a mean of 3.52 and a standard deviation of 0.5. What is the percentage if their CGPA is between 3.33 and 3.61? (5 marks)

[TOTAL 25 MARKS]

Question 5

a. An ultrasonic sensor measures the time interval (μs) of an echo to determine the distance (cm) of an object. The graph below shows the relationship of 6 samples between an echo's time interval (μs) and its distance (cm) of an object.



i. From the scatterplot above, complete the following table. (2 marks)

x	y	x^2	xy
2	293	**(iii)**	**(iv)**
3	337	9	1011
4	**(ii)**	16	1680
5	353	25	1765
(i)	781	49	5467
9	834	81	7506
$\sum x = 30$	$\sum y = 3018$	$\sum x^2 = 184$	$\sum xy = 18015$

Continued.....

ii. Given $SS_{xx} = 34$ and $SS_{xy} = 2925$. If the least square regression line is $\hat{y} = a + bx$, find the value of a and b. (4 marks)

iii. Use the least square regression line to estimate the distance (cm) of an object if its echo's time-interval is 8μs. (2 marks)

iv. Given that $SS_{yy} = 287890$, determine and interpret the coefficient of correlation, r. (3 marks)

b. A recent survey is conducted to investigate that there is a high number of local people visited the Sarawak Cultural Village compared to the international visitor. The result is shown as below.

	Local visitor	International visitor
Sample, x	601	561
Sample size, n	1000	1000

i. Identify the two sample proportions. (2 marks)

ii. State the null and alternative hypotheses. (2 marks)

iii. By using the hypothesis testing of two population proportion, calculate the test statistic, z. (3 marks)

iv. Determine the *p-value*. (3 marks)

c. A statistician claims that university students spend an average of RM800 on food with a standard deviation of RM350 monthly. To test the claim, 42 students were chosen at random and found that the average of their food expenditure is RM700.

i. State the null and alternative hypotheses for the above situation. (2 marks)

ii. Verify the claim using one mean test at 5% significance level. (7 marks)

[TOTAL 30 MARKS]

End of page.

APPENDIX – KEY FORMULAE

- **Mean for ungrouped data**

$$\bar{x} = \frac{\sum x_i}{n} \qquad or \qquad \mu = \frac{\sum x_i}{N}$$

- **Mean for grouped data**

$$\bar{x} = \frac{\sum m_i f_i}{\sum f_i} \qquad or \qquad \mu = \frac{\sum m_i f_i}{\sum f_i}$$

Where m is the midpoint and f is the frequency of the class

- **Variance for ungrouped data**

$$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{\sum (x_i)^2}{n}\right] \qquad or \qquad \sigma^2 = \frac{1}{N}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{N}\right]$$

- **Variance for grouped data**

$$s^2 = \frac{\sum m_i^2 f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}}{\left(\sum f_i\right) - 1} \qquad or \qquad \sigma^2 = \frac{\sum m_i^2 f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}}{\left(\sum f_i\right)}$$

Where m is the midpoint and f is the frequency of the class

DISCRETE RANDOM VARIABLE, BINOMIAL DISTRIBUTION, AND POISSON DISTRIBUTION

- Mean of a discrete random variable x : $\mu = \sum xP(X = x)$
- Standard deviation of a discrete random variable x : $\sigma = \sqrt{\sum x^2 P(X = x) - \mu^2}$
- Binomial probability formula : $P(X = x) = \binom{n}{x} p^x q^{n-x}$

- Poisson probability formula : $P(X=x)=\dfrac{e^{-\lambda}\lambda^{x}}{x!}$

- Key formula for Binomial Probability Distribution :
  - If $p \le 0.5$
    i. $P(X=r)=B(r)-B(r-1)$
    ii. $P(X\ge r)=1-B(r-1)$
    iii. $P(X\le r)=B(r)$
    iv. $P(X>r)=1-B(r)$
    v. $P(X<r)=B(r-1)$
    vi. $P(a\le X\le b)=B(b)-B(a-1)$

  - If $p > 0.5$
    i. $P(X=r)=B(n-r)-B(n-r-1)$
    ii. $P(X\ge r)=B(n-r)$
    iii. $P(X\le r)=1-B(n-r-1)$
    iv. $P(a\le X\le b)=B(n-a)-B(n-b-1)$

- Key formula for Poisson Probability Distribution :
  - If $\mu=\lambda$
    i. $P(X=r)=Poi(r)-Poi(r-1)$
    ii. $P(X\ge r)=1-Poi(r-1)$
    iii. $P(X\le r)=Poi(r)$
    iv. $P(X>r)=1-Poi(r)$
    v. $P(X<r)=Poi(r-1)$
    vi. $P(a\le X\le b)=Poi(b)-Poi(a-1)$

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $z=\dfrac{x-\mu}{\sigma}$

REGRESSION

- Coefficient of correlation, $r = \dfrac{SS_{XY}}{\sqrt{SS_{XX}SS_{YY}}}$ where,

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \quad , \quad SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad , \quad SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

- For the **least square regression**, $\hat{y} = a + bx$

where, $b = \dfrac{SS_{xy}}{SS_{xx}}$, $a = \bar{y} - b\bar{x}$

HYPOTHESIS TESTING

- **ONE population**

Mean:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Proportion:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

- **TWO population**

Mean:

$$z = \frac{\left(\overline{x_1} - \overline{x_2}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^{\,2}}{n_1} + \dfrac{\sigma_2^{\,2}}{n_2}}}$$

Proportion:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \textbf{where} \qquad \hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$$